

Perspectives and Challenges for DNA Microarrays in Drug Discovery and Development

Hans Gmuender

GeneData AG, Basel, Switzerland

DNA microarrays, an indispensable part of the future system of genomics, proteomics, and integrated biology, allow the measurement of the expression of thousands of genes simultaneously. Functional gene expression studies provide insights into the transcription pattern changes that are the consequence of development processes, physiological stimuli, and diseases; they also permit an understanding of functional pathways and the regulation of cellular processes. DNA microarrays can also help in drug discovery and development to confirm the validity of existing targets, to identify new targets, and to detect the consequences or the genetic cause of diseases. Further applications are the use of microarrays in diagnostics and toxicogenomics. However, some problems that arise by using this technique are not yet completely resolved. The reproducibility of experiments, the sensitivity to detect gene expression at low levels, and the conformity to standardize gene expression analysis need further improvements. Therefore, not only the technology itself has to be improved but also genomics has to be strongly coupled to bioinformatics to provide tools for the statistical interpretation and visualization of the data and to set standards so that the data can be meaningfully combined and mined.

TRADITIONAL DRUG DISCOVERY AND DEVELOPMENT

Drug discovery in pharmaceutical companies is traditionally performed by searching for and developing a compound (e.g., an inhibitor) that interferes with an enzyme or a receptor involved in a pathophysiological process. In antibiotic research, the target for a drug is usually an enzyme that, ideally, is essential for the bacteria but not present in humans. Because of the progress made in molecular biology, the enzyme or receptor can frequently be purified, and the kinetic parameters of the enzyme activity, preferably the rate-limiting step in the pathway, are characterized. Structurally diverse small molecules from chemical libraries, combinatorial chemistry, and supernatants of natural products can then be

tested by high-throughput screening for their inhibitory activities against the purified enzyme. If sufficient information about the mechanism of action and the structure of the enzyme is available, then more specific classes of small molecules can be tested and a chemical library can be reduced to more promising drug candidates, which may accelerate the drug development process. In some cases, no suitable leads are found by the process described above, but because the identification of an appropriate lead structure is a vital step for any pharmaceutical research project, new approaches have been developed to find lead compounds. A rational approach to generate lead structures, for example, is to use the detailed 3-D structural information of a substrate binding site on an enzyme to perform initially an *in silico* screen for potential inhibitors (1). Afterward, medicinal chemists try to optimize the lead compounds using their experience and knowledge from earlier and perhaps quite different projects. The goal of this optimization process is to endow the compounds with desirable properties such as bioavailability, high specificity for the target enzyme, capacity to penetrate into cells, no adverse drug effects, etc. For quite a long time, this approach, or comparable biochemistry-based approaches, was successful and has provided many effective antibiotics and drugs for a variety of diseases.

CHALLENGES FOR PHARMACEUTICAL COMPANIES

Today pharmaceutical companies face more competitive challenges than ever before. Changes and strengthening government regulations, increases in costs, and market pressures force pharmaceutical companies to streamline their research and development processes. Various reports on future challenges for pharmaceutical companies come to similar conclusions. (i) Higher investment in research and development in combination with reduced forecasts on future sales leads to lower than expected turnover rates. (ii) At current attrition rates, the increase in the number of compounds will not be sufficient to produce the number of drugs required to maintain current growth rates. (iii) The integration of new technologies such as genomics and proteomics into the drug discovery process increases the number of biological targets but does not necessarily shorten the time to develop a specific drug.

Pharmaceutical companies expect that increasing the number of new chemical entities that go into development would automatically increase the output of successful new drug applications. In fact, there are more drugs in the pipeline for preclinical testing, indicating that the pharmaceutical industry could eliminate a certain "innovation deficit", but, unfortunately, the costs for research and development are soaring dramatically. Therefore, just pushing more drugs down the pipeline will not sort them out. The quality of the drugs has to be improved by changing the drug development process by being much more selective at earlier stages in the research and devel-

opment process to detect, with greater probability, poor drug candidates.

IMPROVING THE RESEARCH AND DEVELOPMENT PROCESSES

During the last two decades, the pharmaceutical industry invested heavily in the life science area. As a consequence, new technologies and methodologies have been developed that can help to streamline the research and development processes. High-throughput screening enables researchers to quickly identify promising candidates for further development out of large compound libraries. Combinatorial chemistry enables chemists to synthesize the large amounts of diverse compound libraries required by biologists. The development and constant improvement of molecular biology (e.g., gene cloning techniques, in vitro expression technology, and site-directed mutagenesis) transformed drug discovery drastically; however, two major hurdles remained. The number of potential targets was limited by the number of cloned genes. Target identification and validation, which require a detailed understanding of a given disease, were time consuming and laborious. This identification barrier has almost been eliminated because an increasing number of human genes are available as targets. In addition, many other organisms, including numerous bacterial genomes, have been sequenced. The validation barrier will soon be surmounted because techniques allowing the matching of human genes to diseases on a whole genome scale have been established. Such approaches include genome scans in disease linkage and association studies, as well as gene expression profiling using DNA microarrays and proteomics. Two-dimensional gel electrophoresis and protein microarrays allow the comparison of changes at the protein level through monitoring at the level of translation and the testing of many enzyme activities simultaneously. DNA microarrays enable the monitoring of the expression levels of thousands of genes simultaneously and can be used for both genotyping and for the determination of mRNA levels.

GENOMICS, PHARMACOGENOMICS, AND DNA MICROARRAYS

Genomics attempts to look, in a much more global sense than traditional genetics, at all the genes used to assemble a living creature as a dynamic system and to determine how these genes interact and influence biological pathways and cell physiology. The goal of genomics is to describe primary DNA sequence in coding and regulatory regions, polymorphic variations within species or subgroups [single-nucleotide polymorphisms (SNPs)], and temporary changes in gene expression patterns during development as a consequence of physiological response or disease and the subcellular localization and intermolecular interaction of protein products.

Genomics has contributed to pharmacology in the area of pharmacogenetics, which describes the influence of genetic factors on the action of drugs (e.g., which drugs work best on which patients) and the genetic basis of susceptibility to adverse reactions of drugs. Pharmacogenomics implies the application of genomic technologies to drug discovery and development.

DNA microarrays establish such global views at the gene expression level by providing a systematic way to examine individual DNA and RNA variation. Functional gene expression studies provide insights into the transcription of RNAs during developmental processes, throughout the cell cycle, and as a consequence of physiological responses or diseases, as well as an understanding of functional pathways, metabolic pathways, and the regulation of cellular processes.

GENE EXPRESSION PATTERNS IN NORMAL CELLS AND TISSUES

The simplified idea is that the genes that are active and therefore expressed in a cell determine which proteins will be made. For instance, these proteins, enzymes, receptors, or ion channels, in turn, not only determine the structure and function of a cell but also strongly influence how the cell will respond to its physiological environment or to a drug. Gene expression may provide indirect information about the function of its corresponding protein; therefore, gene expression analysis can provide insights into temporally induced changes in cells and may indicate whether a gene is of possible significance in a biological process (e.g., during development, during the cell cycle, for maintenance of certain cell functions, after a physiological stimulus, etc.). Gene expression analysis allows the description of complex metabolic pathways; as a consequence of the increasing knowledge about genes and proteins, it will be possible to define cell-, time-, and stimulus-specific patterns. Gene expression analysis also enables the identification and classification of presently uncharacterized genes or new functions for known genes. However, expression data have their limitations because mRNA levels may not always reflect true protein levels, which can be regulated differently.

There are some points to consider when performing gene expression analysis. The expression pattern is only as specific as the source of mRNA, and tissues may be composed of many different cell types that can be difficult to separate. For reasonable gene expression comparisons, it may be very useful to use histological data to describe as exactly as possible the composition of the tissue, the grade of, for instance, a disease, and to dissect as exactly as possible the desired cells; otherwise, it will be very difficult to obtain acceptable reproducibility (5). The use of cell cultures may help to alleviate certain of these problems, and subsequent gene expression studies may result in much clearer patterns, but the disadvantage of cultured cells is that they do not represent a normal physiological situation. Although these cells are derivatives of specific natur-

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

al cell types, they are not in strong contact with other cells, as in complex tissues. Since they are stimulated to grow in culture, they may harbor possible mutations.

Arrays to monitor gene expression should allow the measurement of transcripts across several orders of magnitude to levels of just a few copies per cell. However, to be in or to increase the linear range, one should ideally reduce the amount of cDNA for high-abundance genes and increase the quantity for low-abundance genes to both avoid saturation and to measure signals clearly above background. A better solution would be to systematically change the hybridization conditions and/or to mirror the time course of the hybridization step as it is used, for example, for surface plasmon resonance spectroscopy. Obviously, a major challenge is the detection of genes that show a fluorescent signal just slightly above background and to decide whether the signal is considered noise or is a sufficiently significant signal. However, it is also very important to detect these low-abundance genes for a complete global view. These genes are often the most interesting ones because their products may be at the beginning of a cascade and may regulate the expression of many other genes. Therefore, it is essential to consider very carefully the number of experiments, the reproducibility of the results, and analysis of the data with adequate statistical tools (3,6). Related to this is the problem of specific gene expression pattern characterization. As mentioned above, it is the goal of genomics to define gene expression patterns; therefore, it is essential to detect even subtle differences between patterns. Finally, a prerequisite for successful comparisons is to set up common protocols, assay descriptions, and analysis methods since standardization is an issue of great importance.

GENE EXPRESSION PATTERNS IN PATHOGENS

The small size of bacterial genomes allows the design of a single microarray where every gene from a given bacterium is represented. This opens many new possibilities for the microbiologist and for antibiotic research. Microarray technology will allow the identification of genes that are turned on *in vitro* but not at the site of infection *in vivo* and those genes that are only turned on during infection. Such genes may encode virulence determinants that are regulated by environmental signals. An interesting question that can be answered by these experiments is how bacteria can change gene expression patterns in changing environments. At the current sensitivity level of the technology, the number of cells necessary to isolate enough RNA and the separation of bacterial cells from eukaryotic cells are probably, for many studies, the limiting steps if uncultured bacteria are used. Similar approaches could be used to study viral gene expression during the time course of acute infection or during latency, and microarrays can also be used to study the response of the host to the pathogens (host-defense). The problems described above—reproducibility, sensitivity,

and standardization—are similar to those of gene expression studies with cells or whole tissues. Despite these problems, DNA microarrays will open enormous possibilities to gain global views about the functional genomics of pathogens.

DIFFERENTIAL GENE EXPRESSION PATTERNS IN DISEASES

Genomics, proteomics, and pharmacogenomics promise novel therapies for some diseases and preventative treatments that could reduce disease incidence and severity. It may be discovered, for instance, why certain genetic profiles react differently to a drug or lead to the different metabolism of a drug. A genotype indicative of a certain disease or with a predisposition for a disease may need a preventative therapy or simple disease monitoring. This development will greatly enlarge the market and shift healthcare towards diagnostics and new pharmaceuticals.

Gene Expression as the Consequence of a Disease

The differences that can be observed between gene expression patterns in healthy and diseased cells provide valuable information about the genes and proteins that are involved in disease. Therefore, one of the most attractive applications of microarrays is the study of differential gene expression in disease. An exciting application is probably the characterization of gene expression profiles that are the consequence of a specific disease. This application may lead to more detailed insights into the disease-associated physiological processes and may confirm whether a currently used target is a reliable target. One can imagine that many drugs may suppress the disease, but, for symptom alleviation, there may exist more appropriate targets, enzymes that are more specific for a given cell type, more sensitive, easier to inhibit, essential for the rate-limiting step in a pathway, etc. Therefore, gene expression analysis of diseased cells may help to identify or define new targets that can be regulated to prevent the manifestation of the disease.

Animal disease models offer some advantages to circumvent some of the problems that arise when studying human diseases. Disease models in inbred animals are attractive and very often useful, as they are more amenable to well-controlled experimental design. Gene expression in model organisms has been commonly performed using lower-throughput approaches, such as differential display PCR, and the use of DNA microarray analysis will allow a much higher throughput. However, the relevance to human disease needs to be carefully evaluated, as gene sequence and gene function are not always conserved between species and the manifestation of diseases can vary. Nevertheless, detailed profiling of gene expression in disease model systems will allow additional insights into the physiological processes in humans, which may be critical for the discovery and validation of therapeutic targets. Applications will include the comparison of

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

gene expression patterns not only in normal models but also in models where certain genes are overexpressed, mutated, knocked out, or where the animal gene is replaced by the corresponding human gene.

Mutations, SNPs, Haplotypes—The Causes of a Disease

The genes in human DNA are and will be mapped and decoded, which also means that we will begin to understand which genes are related to what disease. It will be possible, for instance, to test whole populations and identify those who have a genetic risk for certain diseases (for a detailed review, see Reference 2). Targeting gene products that have been identified as the origin of a specific disease can open new ways to achieve disease modification, and the detection of SNPs, haplotypes, and mutations offers many possibilities to develop new therapies and new preventive pharmaceuticals. Treatments for diseases with a genetic basis may finally include gene therapy, protein therapy, and small molecule therapy, which are all used to counter the effects of altered or missing proteins.

Gene Expression and Diagnostics

Gene expression combined with results from SNPs and mutation analysis has applications within diagnostics, too. Biopsy samples or blood cells from patients with a certain disease can be tested to identify the specific genetic cause of their disease, or the samples can be used to determine disease-specific gene expression patterns as the consequence of a disease. The analysis will also show the probable response of the patients to a drug therapy and how they will metabolize the drug. It may then be possible to offer drugs tailored to individual genetic needs. This will change the clinician's traditional trial-and-error search for an effective cure with no or minor side effects. Treatment will be specific to each genetic profile, and patients will receive the right treatment the first time.

GENE EXPRESSION IN RESPONSE TO DRUG TREATMENTS

Drugs generally exert their therapeutic effects by interacting with their targets in such a way that the normal function of the cell is restored or that the effect compensates for any abnormalities. A drug may be designed for a particular enzyme, being optimized *in vitro* but with unknown *in vivo* effects. With conventional methods, it is impossible to completely explore the mode of action of a drug in relation to gene expression patterns and pathways. Therefore, expression analysis can also be used as a highly sensitive indicator for the activity and effects of a drug. In addition, investigating the mechanism of a drug may also help to identify other, better targets. In summary, comparing healthy, diseased, and

drug-treated cells may lead to a much better understanding of the mode of action of a drug.

A few questions that most likely are similar for all drug interaction experiments are raised during the design of such experiments (4). Evidently, a rapid and careful isolation of the cells and of the RNA is a very critical step. It seems to be straightforward to cultivate and collect cells that are in a similar growth phase (e.g., within the exponential growth phase), but one has to keep in mind that the culture always represents a distribution of cells in different growth phases. Interesting in this context are experiments that arrest cells in a well-defined growth phase before analyzing gene expression. Adding a drug raises the question of the concentration. The drug concentration may be very critical because one would probably like to determine the primary response to a drug application. Once the physiology of a cell is disturbed by a drug application, only a few genes may be up- or down-regulated as a consequence of an initial response. On the other hand, treating cells with higher drug concentrations may help to analyze and understand more complex response patterns. Another critical issue is the time point at which a change in the gene expression pattern can be expected after drug application. The handling of this problem may be rather guided by technical limitations such as how fast the cells can be collected after the application of the drug. It is also obvious that it will usually be impracticable to test many possibilities to find out the optimal experimental conditions, but such considerations may be critical for the experimental design and whenever possible should be taken into consideration. Analyzing the expression pattern raises another question: What is a relevant drug response, and what can be considered a significant difference compared to control cells or to cells treated with another drug? For instance, it is known that quinolones vary slightly in their mode of action, and, therefore, one would like to see not only the difference to untreated cells but also the very subtle differences between different quinolones.

An important and interesting application for determining gene expression patterns in the presence of a drug is the testing of compounds with an unknown mode of action. Some compounds are known that show clearly an antibiotic activity, but their targets, and therefore their mode of action, are unknown. Comparing such compounds with existing drug profiles could help to classify unknown drugs and maybe to identify new targets. Based on results from whole-cell screening assays, one could likewise be interested in profiling compounds from combinatorial chemistry or from natural products. Gene expression analysis would also be a powerful tool to test their mode of action. However, the first problem arises in such applications at an even earlier stage—how pure must a compound be to insure that an observed effect is in fact due to the application of a compound and not due to some impurities? Also, compounds or mixtures of compounds, which are or contain so-called “frequent hitters”, should be excluded from microarray experiments because their mode of action is likely unspecific.

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

GENE EXPRESSION IN TOXICOLOGY AND TOXICOGENOMICS

Another rapidly emerging commercial application for gene expression analysis is the field of predictive toxicology and toxicogenomics. Such analysis will complement traditional pathology and toxicity studies, especially in cases where the drug itself is difficult to track by conventional detection methods or when its clinical outcome is a slow process. Matching a gene expression profile of a new drug candidate with profiles of known compounds, which, for instance, showed adverse drug effects, will indicate at a very early stage whether it is worth developing the compound or whether it will most likely have undesired side effects. Gene expression analysis will be used to describe and predict adverse drug effects. Databases produced through large-scale gene expression analysis will provide profiles for many drugs and will help to reduce the substantial amount of time invested in research and discovery experiments. More significantly, such comparisons will reduce the costs for expensive late-stage drug development processes and allow more rapid stop-go decisions during early drug development stages.

CONCLUSIONS AND PERSPECTIVES

The life science industry can look forward to the fact that as more genomes are mapped there will be a wealth of information available from gene expression analysis. The first global functional genomic maps will begin to shed new light on transcription, translation, subcellular localization, mutant phenotype, genotypic variation, replication, recombination, protein-DNA interactions, cell synthesis, and more. The number and variety of microarray applications appear to be unlimited and are an indispensable part of genomics, pharmacogenomics, and proteomics. The techniques and sensitivity will continue to advance for many diverse areas of science.

However, analysis through gene expression studies also creates enormous informatics challenges. This issue is as fundamentally important as the advances in the technology itself. To exploit the commercial potential of microarrays, and more generally of genomics and pharmacogenomics, pharmaceutical companies have to manage data on a very large and complex scale. This requires a large investment in informatics for drug discovery and development processes, and new technologies are required to collate, manage, analyze, and distribute the vast amounts of data. The data can no longer be handled by deductive exploration but needs an inductive analysis, usually summarized in the literature as "Data Mining". Bioinformatics will influence the design of experiments because of the need to be large-scale, global, and system-wide. Informatics must provide tools for the statistical interpretation of the data sets, for the visualization of gene expression patterns, and should influence the settings of common standards so that investigators can meaningfully combine, mine, and interpret data.

REFERENCES

1. **Boehm, H.-J., M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne et al.** 2000. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* *43*:2664-2674.
2. **Chakravarti, A.** 1999. Population genetics—making sense out of sequence. *Nat. Genet. Suppl.* *21*:56-60.
3. **Duggang, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent.** 1999. Expression profiling using cDNA microarrays. *Nat. Genet. Suppl.* *21*:10-14.
4. **Gmuender, H., K. Kuratli, K. Di Padova, C.P. Gray, W. Keck, and S. Evers.** 2001. Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res.* *11*:28-42.
5. **Heller, R.A., J. Allard, Z. Fengrong, C. Lock, S. Wilson, P. Klonowski, H. Gmuender, H. Van Wart, and R. Booth.** 1999. Gene chips and microarrays: applications in disease profiles, drug target discovery, drug action and toxicity, p. 187-202. *In* M. Schena (Ed.), *DNA Microarrays: A Practical Approach*. Oxford University Press, New York.
6. **Schadt, E.E., C. Li, C. Su, and W.H. Wong.** 2000. Analyzing high-density oligonucleotide gene expression array data. *J. Cell. Biochem.* *80*:192-202.

Feedback and suggestions for contributions to the "Drug Discovery and Genomic Technologies" section are welcomed by the Scientific Editor, Dr. James Ellingboe (ellingboe@BioTechniques.com).

**For reprints of this or any other article,
contact Reprints@BioTechniques.com**